

R DAY 2018: UMA API EXTENSÍVEL PARA QUEBRAR CAPTCHAS

C. Lente

2018-05-22

Decryptr

Motivação

- O `decryptr` nasceu com o objetivo de facilitar a quebra de captchas textuais de serviços públicos
- Muitas vezes os dados são públicos, mas não *acessíveis*, impedindo que um cidadão comum analise-os

- Talvez seja possível quebrar o captcha usando apenas transformações na imagem e um OCR padrão, mas isso acarreta dois problemas:
 - O tratamento das imagens dos captchas de cada fonte não pode ser generalizado para outra fonte
 - Transcrição de textos através de OCR não só tem uma taxa de acerto muito pequena, bem como não é escalável

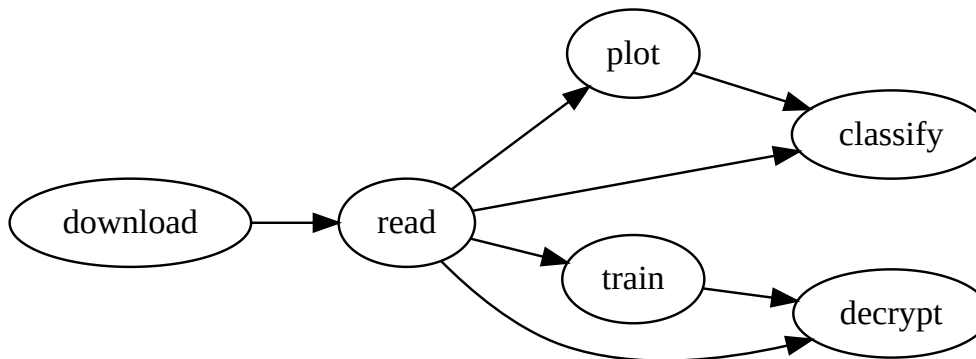
Keras

- Nossa solução para esses problemas foi o Keras, uma API de redes neurais de alto nível que pode rodar em cima de múltiplos backends como TensorFlow
- Usando este framework, Daniel Falbel¹ e Julio Trecenti criaram um uma rede neural que consegue aprender os padrões de famílias de captchas
 - Dado um conjunto de treinamento adequado, é possível quebrá-los com acerto e velocidade impressionantes.
- O pacote já vem com alguns modelos pré-treinados, mas o grande benefício de usar o `decryptr` é que ele permite que qualquer usuário crie seus próprios modelos
 - Os captchas da família precisam ter o mesmo comprimento e a cor de cada letra não pode ser relevante para a quebra do captcha

[1] O Daniel inclusive chegou a contribuir para o desenvolvimento do *port* da biblioteca Keras para o R.

O Pacote

- São essencialmente dois módulos: um para quebrar captchas (`decrypt()`) e um para treinar novos modelos (`train_model()`)
- Temos funções auxiliares como `read_captcha()` para ler e plotar uma imagem, `classify()` para auxiliar na criação de bases de treino e `load_model()` para carregar modelos que o usuário já tenha treinado



Exemplo

```
library(decryptr)

file <- download_captcha("trt", path = "./img")
captcha <- read_captcha(file)
plot(captcha[[1]])

decrypt(file, model = "trt")

files <- download_captcha("trt", n = 3, path = "./img")
new_files <- classify(files, path = "./img")
captchas <- read_captcha(new_files, ans_in_path = TRUE)

model <- train_model(captchas, verbose = FALSE)

decrypt(file, model = model)

microbenchmark::microbenchmark(decrypt = decrypt(captcha, model))
```

Demonstração